Data management workflows for campaigns and model data

Linda Baldewein
Helmholtz Coastal Data Center
Institute of Carbon Cycles
Helmholtz-Zentrum Hereon
Geesthacht, 29.09.2021
Agenda

- Example workflow for campaigns
  - Hereon campaign data workflow
- Example workflow for models
  - Introduction to the Model Data Explorer
Campaign data workflow

Step 1: Campaign planning

Step 2: Digital metadata collection

Step 3a: Mobile data transfer

Step 3b: Laboratory analysis

Step 4: Long-term storage of quality checked (meta-) data

Step 5: Publication of ISO 19115–3:2016 metadata files

Step 6a: DOI registration with WDC PANGAEA

Step 6b: Sample registration (work in progress)

Step 7: Integration in Hereon, national and international data portals

Planning
Field
Laboratory
Cloud
Database
Metadata catalog
DOI
Portal

IGSN

Data workflows for campaigns and model data
Campaign data: Planning

- Where will the campaign take place?
- How many events / stations are planned?
- Create planned event / station list and / or web map
  - Most German research institutions have an ArcGIS Online subscription without you knowing.
  - QGIS as open-source Desktop alternative
Campaign data: Planning

- How will metadata be recorded?
  - DSHIP on larger research vessels
  - Handwritten
  - Survey123 App
- Who will record metadata?
Campaign data: Collecting

Required information
- Metadata of your campaign
  - Ship name
  - Station / Event name
  - Coordinates
  - Sample information
  - Participants

-> Ideally digitally collected using DSHIP or Survey123
Campaign data: Processing and Analysing

- Metadata of your measurement
  - Parameter name
  - Unit
  - Method
  - Quality information
  - Origin of sample
Campaign data: Preserving

- Store results locally
  - Meaningful file names
  - Shared storage with your working group
  - Include metadata
- Prepare for submission
  - Format based on requirements of repository
  - Use template, if available
  - Double check for typos, etc.
- Submit to institution repository (if available)
Campaign data: Publishing and Sharing

- Publish your data
  - DOI registration with PANGAEA
- See details in talk by Flavia Höring
- Publish your samples
  - IGSN
- Your local data curators may help you. Ask them!
Campaign data: Reusing

- Data available in data portals
- Other scientists can find, download and cite your data
Model data: Planning and Collecting

- Which model will you use / develop?
- Gather data
  - Reuse data published by other scientists
Model data: Processing and Analysing

- Metadata of your model
  - Conventions
  - Institution, originator, contact
  - Title
  - Source
  - Creation date
  - Coordinate reference system
  - Dimensions
  - Variables
  - Units
  - Standard names
Model data: Preserving

- Store results locally
  - Meaningful file names
  - Shared storage with your working group
  - Include metadata
- Prepare for submission
  - Format based on requirements of repository
  - Use template, if available
  - Double check for typos, etc.
- General rules for data storage in netCDF files compiled at Hereon

2. General Specifications

Provided here is general information that has nothing directly to do with the netCDF format itself but should, however, be taken into consideration.

2.1. File Names

- NetCDF files are designated with the extension .nc
- The file name begins with a letter or a letter sequence. The letter sequence should allow users to draw conclusions about the type of data found in the file.
- Customary sequences used thus far (e.g., “ctd” for CTD data, “sf” for ScanFish data) will be retained. Specification should still be designated for other devices.
- When dealing with model data, either the model name with the version designation or the ExperimentID should be used in the first segment of the filename.
Model data: Publishing and Sharing

- Publish your data
  - DOI registration with WDC CERA at DKRZ
- Your local data curators may help you. Ask them!
Model data: Reusing

- Other scientists can find, download and cite your data
- Data available in data portals
  - Currently a lot of manual effort by curators
Model Data Explorer

- Central platform to access Model Data
  - 4D Model Data Map
  - Search and filter datasets
  - Compute and compare statistics on the data
  - Download raw data
- Currently in development
  - We need support from future users
  - Join our mailing list mde-dev@listserv.dfn.de
  - Check out the development
Vielen Dank.